

Research article

Design of Phase II cancer trials evaluating survival probabilities

L Douglas Case* and Timothy M Morgan

Address: Department of Public Health Sciences and the Comprehensive Cancer Center of Wake Forest University, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC, USA

Email: L Douglas Case* - dcase@wfubmc.edu; Timothy M Morgan - tomorgan@wfubmc.edu

* Corresponding author

Published: 3 April 2003

Received: 30 September 2002

BMC Medical Research Methodology 2003, **3**:6

Accepted: 3 April 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/6>

© 2003 Case and Morgan; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Phase II cancer studies are undertaken to assess the activity of a new drug or a new treatment regimen. Activity is sometimes defined in terms of a survival probability, a binary outcome such as one-year survival that is derived from a time-to-event variable. Phase II studies are usually designed with an interim analysis so they can be stopped if early results are disappointing. Most designs that allow for an interim look are not appropriate for monitoring survival probabilities since many patients will not have enough follow-up by the time of the interim analysis, thus necessitating an inconvenient suspension of accrual while patients are being followed.

Methods: Two-stage phase II clinical trial designs are developed for evaluating survival probabilities. These designs are compared to fixed sample designs and to existing designs developed to monitor binomial probabilities to illustrate the expected reduction in sample size or study length possible with the use of the proposed designs.

Results: Savings can be realized in both the duration of accrual and the total study length, with the expected savings increasing as the accrual rate decreases. Misspecifying the underlying survival distribution and the accrual rate during the planning phase can adversely influence the operating characteristics of the designs.

Conclusion: Two-stage phase II trials for assessing survival probabilities can be designed that do not require prolonged suspension of patient accrual. These designs are more efficient than single stage designs and more practical than existing two-stage designs developed for binomial outcomes, particularly in trials with slow accrual.

Background

Phase II clinical trials are usually conducted to assess the activity of a new drug or treatment regimen. Activity in many phase II cancer trials is quantified by tumor response, and a treatment is considered successful for an individual patient if his or her tumor burden is reduced by 50% or more. Here, the response rate is the number of patients responding divided by the total number of evalua-

ble patients. In other trials, activity might be quantified by a time to event variable such as remission-free or overall survival, and the outcome might be the proportion of patients remission-free or alive at one or two years. Regimens showing sufficient activity in the phase II setting might be evaluated subsequently in a phase III comparative trial. Phase II trials typically test the null hypothesis $H_0: p \leq p_0$ that the true response rate of the treatment un-

der consideration is less than some level (p_0) that would be deemed too low for further consideration. Studies are designed so that the probability of falsely rejecting H_0 (i.e., considering the treatment worthy of further investigation when in fact it is not) is α and the probability of rejecting H_0 when $p = p_1$ (i.e., considering the drug worthy of further investigation when in fact it is) is $1 - \beta$.

Due to ethical concerns, interim analyses are done in phase II trials to ensure that patients are not receiving a treatment that is clearly inferior to other available options. Usually these trials are not stopped early when a treatment has shown better than expected activity because there is no ethical dilemma, and there is usually interest in obtaining better estimates of the treatment's activity. On the other hand, if the treatment is performing poorly, physicians want to explore other options for subsequent patients. Numerous frequentist and Bayesian multi-stage phase II designs have been developed for monitoring binary outcomes [see, for example, [1–10]]. Simon [4] developed two-stage plans that minimized the maximum expected sample size and plans that minimized the expected sample size under the null hypothesis. He felt one interim analysis was usually sufficient since it is frequently impractical to analyze the data multiple times, and two-stage designs realize much of the savings possible with sequential or group sequential designs [11]. For example, Chen [7] extended Simon's designs to three stages and found a mean reduction in sample size of only 10% with the addition of an extra stage.

When the response of interest is based on a time to event outcome (e.g., one-year survival), censoring becomes an issue during the interim analysis, as it is unclear how subjects without sufficient follow-up should be handled. The simple proportion of all subjects surviving the required time is a biased estimate of the survival probability if some subjects have incomplete follow-up, and restricting the estimate to those subjects followed for the required time results in an inefficient estimate. On the other hand, suspending accrual while all the subjects are followed the required length of time is impractical. Long trial suspensions can ruin a trial's momentum, increase study length, and increase costs, and it is potentially unclear how to treat new subjects during the suspension.

Example

In 1998, investigators in the Comprehensive Cancer Center of Wake Forest University wanted to design a phase II study to assess the activity a new chemo-radiation combination for patients with resectable pancreatic cancer. Pancreatic cancer is the most deadly of all the major cancers. When this trial was planned, 29,000 new cases of pancreatic cancer and 28,900 deaths from the disease were expected in the United States [12]. Patients with resectable

pancreatic cancer have a better prognosis than those with unresectable disease, but the long-term outlook for these patients is still bleak. Investigators felt that Gemcitabine, a radiation sensitizer that had shown activity in pancreatic cancer [13–16], combined with external beam radiation, could improve the prognosis of these patients. Since the tumors were resected in these patients, objective tumor response was not a possible outcome, and one-year survival was chosen as a clinically meaningful outcome. The investigators decided that the treatment would be considered unsuccessful if the one-year survival was 35% or less, and it would be considered active enough to pursue further if the one-year survival was 50% or greater.

The fixed sample size required for a single-stage study to test this hypothesis, based on an exact binomial test, would be 72. If 31 or more of the patients live longer than one year, the null hypothesis would be rejected. This design has type I and II errors of .096 and .097, respectively. Assuming accrual of 24 patients per year (for a duration of accrual of 3 years), each patient would be followed until failure or for a maximum of one year and the single analysis would be done at 4 years.

Now suppose we would like to do one interim analysis during the course of the study. As mentioned previously, Simon's [4] two-stage designs could be used to accomplish this goal. Thirty-four patients would be needed during the first stage. Each of these patients would be followed until failure or for a maximum of one year. After each of the original 34 patients has been followed for one year, the interim analysis would be done. If 13 or more patients live one year, the study would continue until a total of 81 patients had been accrued. All patients would then be followed until failure or for one year. If 34 or more patients live one year then the null hypothesis would be rejected. This design has a 59% chance of stopping at the interim analysis under the null hypothesis. Characteristics of this design are shown in Table 1. The expected sample size under H_0 is 53.2 (or 74% of the fixed sample value). The expected and maximum study lengths are 3.63 and 5.38 years, respectively, (or 91% and 134% of the fixed sample values). As mentioned earlier, the problem with this design is that accrual is suspended for one year while those initially accrued are followed to determine their status. This trial suspension is inconvenient and can lead to a long study (in this example the maximum study length is 5.38 years compared to 4 years for the fixed sample design). These problems are exacerbated for longer suspensions.

One might be tempted to use Simon's design, but to accrue additional patients while the initial 34 are followed. Characteristics for this design are also shown in Table 1. For our example, 24 additional patients would be accrued

Table 1: Characteristics of Single Stage, Simon Two-Stage, and Optimal Two-Stage Designs *

Design	ESS (%)	MDA (%)	ETSL (%)	MTSL (%)
Single Stage	72.0 (100)	3.00 (100)	4.00 (100)	4.00 (100)
Simon – no interim accrual	53.2 (74)	3.38 (112)	3.63 (91)	5.38 (134)
Simon – interim accrual	67.4 (94)	3.38 (112)	3.22 (80)	4.38 (109)
Proposed – minimize ETSL	63.5 (88)	3.44 (115)	3.00 (75)	4.44 (111)
Proposed – minimize EDA	62.0 (86)	3.27 (109)	3.08 (77)	4.27 (107)

* ESS is the expected sample size; MDA is the maximum duration of accrual; ETSL and MTSL are the expected and maximum total study length, respectively. Expected values are calculated under the null hypothesis; percentages are calculated relative to the single stage values.

while the first 34 are being followed, and the trial would have an expected sample size under H_0 of 67.4 (or 94% of the fixed sample value). The expected and maximum study lengths would be 3.22 and 4.38 years, respectively (or 80% and 109% of the fixed sample values). One sees that accruing additional patients while the initial patients are followed results in a greater expected sample size, but shorter expected and maximum total study lengths. The two main problems with this design are 1) information from patients who have been followed less than one year during the interim analysis is ignored, and 2) information collected from the additional patients is potentially never used. Additionally, if accrual is fast, the total required recruitment could be completed by the time of the interim review, possibly resulting in more patients than necessary being treated with, what turns out to be, an ineffective regimen. Note that Herndon [9] proposes a hybrid phase II design that allows for interim accrual but uses information collected on all patients by delaying the decision to stop the trial until all data are reviewed.

Several researchers have proposed using the Kaplan-Meier [17] or Nelson-Aalen [18] estimates of the survival probability during interim analyses to account for the information available from those with partial follow-up without necessitating trial suspension. Jennison and Turnbull [19] give an example of assessing the effect of a drug on the mother to infant HIV transmission rate among HIV infected breast-feeding mothers, where the proportion of infants infection-free at two years would be a meaningful outcome. They suggest monitoring such a trial using Kaplan-Meier [17] estimates and the spending function approach originally described by Lan and DeMets [20]. Lin et al [21] discuss the design of a study in young children with Wilms cancer for which two-year relapse-free survival is the primary outcome, and Nelson-Aalen [18] estimates are used during the interim analyses.

In the next section, we apply results presented by Lin et al [21] to show how to design efficient phase II studies for monitoring survival probabilities without the aforementioned drawbacks. We develop designs that minimize ei-

ther the expected duration of accrual or the expected total study length under H_0 . We choose optimal designs under the null hypothesis since we want to minimize the number of patients treated with an ineffective regimen. Numerical results are presented to illustrate the effect of different choices for the design parameters, including the type I and II errors and the duration of accrual (or, alternatively, the accrual rate). We illustrate these methods in detail for our particular trial and follow with a discussion of possible limitations and extensions.

Methods

Lin et al [21] derived the asymptotic joint distribution of the Nelson-Aalen [18] estimates of survival calculated at different calendar times during a study and applied the results to longitudinally monitor a National Wilms Tumor Study Group protocol. A brief summary of their relevant results is given below. Following their notation, assume n patients are accrued to a trial at times Y_1, \dots, Y_n . Let T_1, \dots, T_n denote the failure times and C_1, \dots, C_n the censoring times since study entry. At time t , we observe for each individual either a failure time or a censoring time and an indicator specifying which. That is, we observe the time $X_i(t) = \min(T_i, C_i, \max(0, t - Y_i))$ and the failure indicator $\Delta_i(t) = I\{T_i \leq \min(C_i, \max(0, t - Y_i))\}$. Let $\hat{S}(x; t) = e^{-\hat{\Lambda}(x; t)}$ denote the estimate of x -year survival at time t , based on the Nelson-Aalen [18] estimate of the cumulative hazard function, $\Lambda(x; t)$:

$$\hat{\Lambda}(x; t) = \sum_{\{i: X_i(t) \leq x\}} \frac{\Delta_i(t)}{R_i(t)} \text{ with } R_i(t) = \sum_{j=1}^n I\{X_j(t) \geq X_i(t)\}.$$

The process, $\frac{1}{n^2} \{ \hat{\Lambda}(x; t) - \Lambda(x) \}$, $x < t$, is asymptotically Gaussian, which, assuming constant accrual and no loss to follow-up, has variance function

$$\sigma^2(x; t) = \int_0^x \frac{\lambda(u) du}{S(u) \min[(t-u)/MDA, 1]}, \tag{1}$$

where $\lambda(u)$ is the hazard function and MDA is the maximum duration of accrual. Lin et al [21] recommend assessing the hypothesis $H_0: S(x) = S_0(x)$ at time t using the asymptotically standard normal test statistic

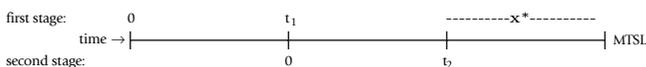
$$Z(x; t) = \frac{n^{\frac{1}{2}} \{ \ln \hat{\Lambda}(x; t) - \ln \hat{\Lambda}_0(x) \} \Lambda(x; t)}{\{ \sigma^2(x; t) \}^{\frac{1}{2}}}, \tag{2}$$

where

$$\hat{\sigma}^2(x; t) = \sum_{\{i: X_i(t) \leq x\}} \frac{\Delta_i(t)}{R_i^2(t)/n}. \tag{3}$$

Let $I(x; t_i)$ denote the information available for estimating $S(x)$ at time t_i . The joint distribution of $Z(x; t)$ calculated over the course of the study is multivariate normal with correlations given by $\rho = \sqrt{I(x, t_i) / I(x, t_j)}$, where $t_i \leq t_j$.

Now consider the design of a two-stage phase II trial for testing $H_0: S(x^*) = S_0(x^*)$, where x^* denotes the survival time of interest and $S(\cdot)$ denotes the survival function which can be estimated as described above (or by using the Kaplan-Meier estimator). As illustrated below, let t_1 and t_2 denote the duration of the first and second accrual periods, and let $MTSL$ denote the maximum total study length ($MTSL = t_1 + t_2 + x^*$).



We will use the following notation:

x^* = survival time of interest

t_i = duration of accrual for the i^{th} stage

n_i = sample size accrued during the i^{th} stage; $n = n_1 + n_2$

v = constant rate of accrual

P_s = probability of stopping at t_1

DA = duration of accrual

EDA = expected duration of accrual = $t_1 + (1 - P_s)t_2$, calculated under the null hypothesis

ESS = expected sample size = $n_1 + (1 - P_s)n_2 = vEDA$, calculated under the null hypothesis

MDA = maximum duration of accrual = $t_1 + t_2$

$ETSL$ = expected total study length = $t_1 + (1 - P_s)(t_2 + x^*)$, calculated under the null hypothesis

$MTSL$ = maximum total study length = $t_1 + t_2 + x^*$

I_1 = information available at t_1

I_{max} = information available at $MTSL$

A two-stage design proceeds as follows.

Stage 1. Accrue n_1 patients between time 0 and time t_1 . Each patient will be followed until failure or for x^* years or until time t_1 , whichever is less. Calculate $Z_1(x^*; t_1)$ as given in (2). If $Z_1(x^*; t_1) < C_1$ stop the study and "accept" H_0 ; otherwise, continue to the next stage.

Stage 2. Accrue n_2 additional patients between times t_1 and $t_1 + t_2 (= MDA)$. Follow all patients until failure or for x^* years, calculate $Z_2(x^*; MTSL)$, and reject H_0 if $Z_2(x^*; MTSL) > C_2$.

An interim analysis could be done anytime after x^* years and before time $MTSL$. The expected duration of accrual (EDA) under the null hypothesis is given by $EDA = t_1 + (1 - P_s)t_2$, where $P_s = \Phi(C_1)$ and $\Phi(\cdot)$ denotes the standard normal distribution function. The expected total study length ($ETSL$) under the null hypothesis is given by $ETSL = t_1 + (1 - P_s)(t_2 + x^*)$. The maximum amount of information for estimating $\Lambda(x^*)$, (I_{max}), occurs whenever $t \geq MTSL$. The joint distribution of Z_1 and Z_2 is bivariate normal with correlation $\rho = \sqrt{I_1 / I_{max}}$.

During the design stage, estimates of the information as a function of time can be obtained by making assumptions regarding the expected survival distribution. Once the survival distribution is specified, the expected information at any time ($I(x^*; t) = 1/\sigma^2(x^*; t)$) can be easily obtained by numerically evaluating equation 1. The Weibull distribution is a flexible and simple choice, as the survival distributions are completely specified by the null and alternative survival probabilities for any given shape parameter.

For two-stage phase II designs for survival probabilities, there are four unknowns – n_1 or t_1 , n_2 or t_2 , C_1 , and C_2 – and two constraints (type I and II errors). We assume the accrual rate is fixed. As there are more unknowns than constraints, there will be an infinite number of solutions. We will choose those solutions that minimize either the expected duration of accrual (i.e., expected sample size) or the expected total study length under the null hypothesis. Note that this is the same paradigm as used by Simon [4] in selecting optimal designs for binomial outcomes, where he only allows early acceptance of the null hypothesis and minimizes the expected sample size under the null hypothesis. Specifically, we choose to minimize the *EDA* or the *ETSL* given that $B(C_1, C_2, \rho) = \alpha$ and $B(C_1 - \rho u, C_2 - u, \rho) = 1 - \beta$, where $u = n^{1/2}(\mu - \mu_0)/\sigma$, μ and σ are the mean and standard deviation of the test statistic, and $B(C_1, C_2, \rho)$ denotes the bivariate normal probability that $Z_1 > C_1$ and $Z_2 > C_2$, given a correlation between Z_1 and Z_2 of ρ . Numerical integration of the bivariate normal distribution is accomplished using a double precision Fortran function written by Donnelly [22]. For initially chosen values for C_2 and ρ , values of C_1 and n are found to satisfy the two error constraints. The parameter values that result in an optimal design are found by iterating over C_2 and ρ using a combination golden-section search and parabolic interpolation minimization routine described by Brent [23]. A given choice of n and ρ corresponds to a specific t_1 , which is obtained by solving equation 1 under the null hypothesis. This, in turn, corresponds to another ρ under the alternative hypothesis, obtained by solving equation 1 under the alternative specifications. This latter ρ is used in calculating the sample size (or duration of accrual) that satisfies the type II error constraint. For practical purposes this step is probably unnecessary since the correlation is very similar under the null and alternative hypotheses, even though the information is different at each stage under the two hypotheses. Fortran code implementing this algorithm is available from the first author upon request. Run times vary between 5 and 10 minutes on an IBM compatible Pentium 3 600 MHz PC; the run time can be decreased by decreasing the precision specified in the program.

Results

Application

Consider the example mentioned earlier. That is, suppose we would like to design a phase II study to assess the effectiveness of adjuvant Gemcitabine and external beam radiation for the treatment of patients with resectable pancreatic cancer. The principal outcome measure used to quantify treatment effect will be one-year survival, and we will test the null hypothesis that one-year survival is 35% or less. We desire to have 90% power at an alternative one-year survival of 50% for testing this hypothesis at the 10% one-sided level of significance.

We use the methods described above to develop two-stage designs that minimize either the duration of accrual or the expected total study length under the null hypothesis. An interim analysis could be done anytime after one year until the end of the study. We assume survival follows a Weibull distribution with a shape parameter of one and a scale parameter of 1.0498 (since $S(1) = .35$ under H_0). Additionally, we assume accrual of 24 patients per year. At this point in the development of the study it might be useful to consider the optimal designs for various interim analysis times. Characteristics of some of these designs are shown in Table 2. The design that minimizes the expected total study length has an interim analysis after 2.2 years. At that time, one calculates $Z_1(x^*; t_1)$ as described above. If $Z_1(x^*; t_1) < .375$ then the study is stopped and the null hypothesis 'accepted'. Otherwise, accrual continues for 1.2 additional years, all patients are followed until failure or for one year, and the null hypothesis is rejected if $Z_2(x^*; MTSL) > 1.172$. This design has an expected sample size of 63.5 (88% of the fixed sample value) and an expected total study length of 3.0 years (75% of the fixed sample value). The design that minimizes the expected duration of accrual has an interim analysis after 1.9 years. At that time, one calculates $Z_1(x^*; t_1)$ as described above. If $Z_1(x^*; t_1) < .004$ then the study is stopped and the null hypothesis 'accepted'. Otherwise, accrual continues for 1.4 additional years, all patients are followed until failure or for one year, and the null hypothesis is rejected if $Z_2(x^*; MTSL) > 1.220$. This design has an expected sample size of 62 (86% of the fixed sample value) and an expected study length of 3.1 years (77% of the fixed sample value). The expected and maximum duration of accrual and total study lengths for these designs are also listed in Table 1 for comparison with the single stage and Simon two-stage designs. One notes that both of the optimal designs have a smaller *ETSL* and *EDA* than one would get from using a Simon design with interim accrual. In addition, they have a smaller *ETSL* and *MTSL* than one would get from using a Simon design without interim accrual. Of course, the latter Simon design would have a smaller expected sample size since it is optimal in that respect.

Figure 1 shows results for optimal designs chosen to minimize the *ETSL* under H_0 for a range of interim analysis times. In Figure 1, one sees that the *ETSL* is fairly flat for optimal designs with interim analyses between 1.8 and 2.8 years, so any design chosen with an interim analysis between these two times will be close to optimal. Thus, a more feasible design might be to have the interim analysis at 2.6 years since the expected total study length for this design is close to optimal and the maximum duration of accrual and the maximum study length are smaller than those for the overall optimal design.

Table 2: Optimal (under H_0) two-stage parameters for testing $H_0: S(I) \leq .35$ vs $H_1: S(I) > .35$. Power is 90% at $S(I) = .5$ *

Minimized	t_1	C_1	C_2	I_1/I_{max}	EDA (%)	MDA(%)	ETSL (%)	MTSL (%)
ETSL	1.8	.137	1.164	.31	2.65 (88)	3.71 (124)	3.10 (77)	4.71 (118)
	2.0	.270	1.164	.38	2.63 (88)	3.59 (120)	3.02 (76)	4.59 (115)
	2.2 @	.375	1.172	.46	2.65 (88)	3.44 (115)	3.00 (75)	4.44 (111)
	2.4	.464	1.184	.53	2.70 (90)	3.32 (111)	3.02 (75)	4.32 (108)
	2.6	.550	1.198	.61	2.78 (93)	3.22 (107)	3.07 (77)	4.22 (105)
EDA	1.5	-.313	1.223	.25	2.66 (89)	3.36 (112)	3.28 (82)	4.36 (109)
	1.7	-.125	1.218	.31	2.60 (87)	3.33 (111)	3.15 (79)	4.33 (108)
	1.9 @	.004	1.220	.38	2.58 (86)	3.27 (109)	3.08 (77)	4.27 (107)
	2.1	.109	1.227	.46	2.60 (87)	3.20 (107)	3.06 (76)	4.20 (105)
	2.3	.189	1.237	.53	2.65 (88)	3.13 (104)	3.08 (77)	4.13 (103)

* t_1 is the time of the first interim review; C_1 and C_2 are the cutpoints for acceptance at the first stage and rejection at the second stage; EDA and MDA are the expected and maximum duration of accrual; and ETSL and MTSL are the expected and maximum total study length. Expected values are calculated under the null hypothesis; percentages are calculated relative to the single stage values. @ optimal design

Misspecification of Survival Distributions and Accrual

Although the test used in assessing the survival probabilities is nonparametric, the choice of a particular design during the planning stage of a trial depends on specification of the survival distribution as well as the accrual rate. Misspecifying these parameters can lead to nonoptimal designs with incorrect type I and II errors. For example, consider the Weibull survival distributions shown in Figure 2 with scale parameters equal to those of the exponential distributions (under the null hypothesis in the figure) from the example above and shape parameters ranging from 0.25 (i.e., events happening earlier on relative to the exponential) to 4.0 (i.e., events happening later). Were we to have designed the trial as above assuming exponential survival (shape parameter = 1), and survival actually followed one of the other Weibull distributions, then we would be getting information faster (shape < 1) or slower (shape > 1) than anticipated, resulting in a larger or smaller correlation and increased or decreased α and $1 - \beta$.

Table 3 illustrates the degree to which α and $1 - \beta$ are affected by misspecification of the survival distributions, assuming the design is used as given and not changed before or after the interim review. It can be seen that misspecification of the survival distribution actually has little effect on α and $1 - \beta$, with α ranging from .095 to .106 and $1 - \beta$ ranging from .880 to .918 as the shape parameter varies from 4.0 to 0.25.

Misspecifying the accrual rate has a much greater impact. Clearly, if accrual were slower or faster than expected, then one would have a lesser or greater proportion of the data at the planned interim analysis. Table 4 shows the effect of misspecifying the accrual rate on the operating characteristics of the design that minimizes the expected total study length. Three scenarios are considered. First we consider what would happen if analyses are done at the

planned times, i.e. t_1 and $MTSL$. Under this scenario, the correlation structure is unchanged so there is no effect on the type I error. However, the effect on the power is substantial since the sample size is much smaller or larger than planned. As the actual accrual decreases relative to the anticipated accrual, the power decreases whereas it increases as the actual accrual increases relative to the anticipated accrual. Under the second scenario, the first analysis is conducted at t_1 . The second analysis is conducted one year after t_1 or after n patients are accrued, whichever is later. Here, the correlation structure and the sample sizes are both affected by the accrual rate, and we see that both α and $1 - \beta$ decrease as the actual accrual decreases relative to the anticipated accrual, whereas both increase as the actual accrual increases relative to the anticipated accrual. Under the third scenario, we conduct analyses after n_1 and n patients have been accrued. Here only the correlation structure is affected by the accrual rate. That is, I_1 depends on how long each of the n_1 subjects is followed. A slower accrual means each subject is followed longer, so I_1 and the correlation increase. Under this scenario, the effects of misspecifying the accrual rate on α and $1 - \beta$ are less substantial. Both α and $1 - \beta$ increase as the actual accrual decreases relative to the anticipated accrual, whereas both decrease as the actual accrual increases relative to the anticipated accrual.

General Results

The design parameters (C_1 , C_2 , t_1 , and t_2) and performance (as quantified by EDA and $ETSL$ relative to the fixed sample values, DA and TSL , respectively) of the phase II designs depend on α , β , x^* , and DA (or alternatively, the rate of accrual), and survival under the null and alternative hypotheses. Table 5 provides optimal design parameters (for minimizing EDA or $ETSL$) and design characteristics for $\alpha = \beta = .05$ and $.10$ for our particular design of interest, assuming survival will follow a Weibull

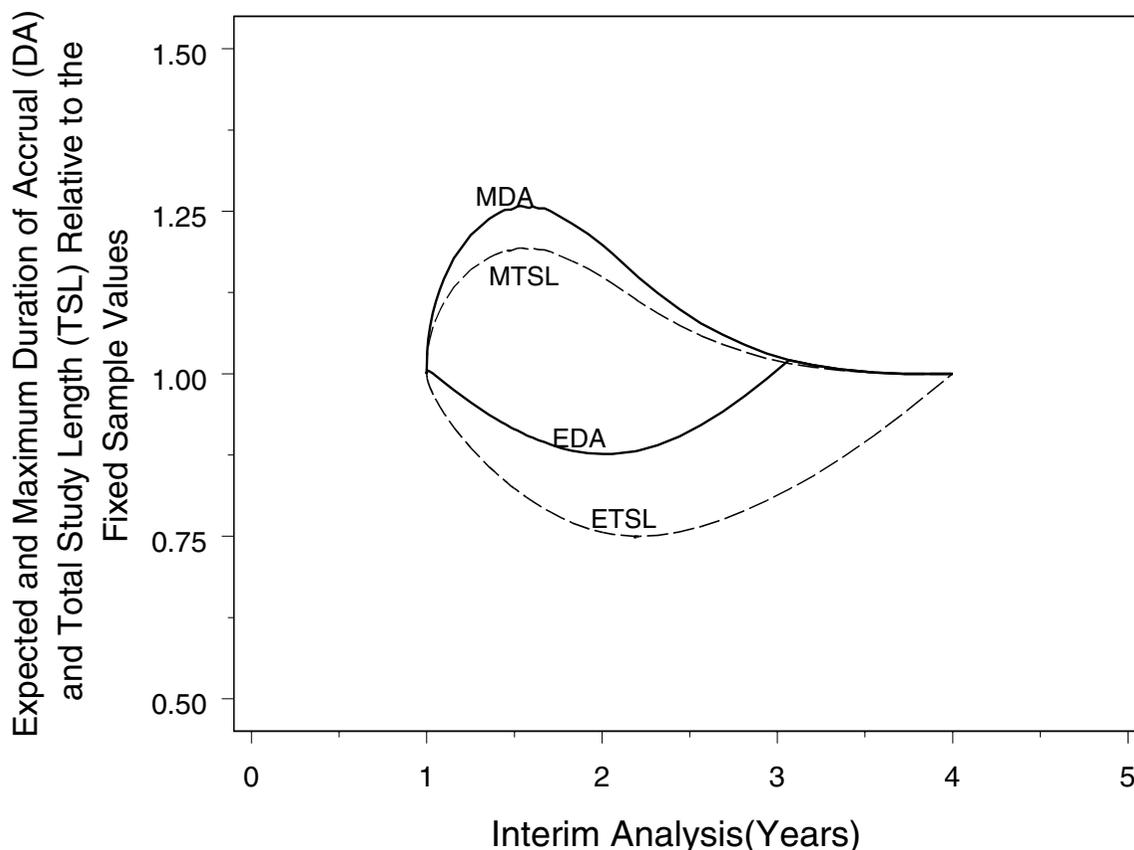


Figure 1
Relative expected and maximum duration of accrual (DA) and total study length (TSL) for designs that minimize the expected total study length (ETSL).

distribution with shape parameter equal to one. Note that the designs depend on the survival time of interest (x^*) and the fixed duration of accrual (DA) in a relative manner. That is, the same design is obtained for assessing 2-year survival when the fixed sample duration of accrual is 4 years as that for assessing 1-year survival when the fixed sample duration of accrual is 2 years. Designs are given for DA/x^* ranging from 1.5 (e.g., $x^* = 2$ and $DA = 3$) to 4 (e.g., $x^* = 1$ and $DA = 4$). Several observations are obvious. As the duration of accrual decreases (accrual rate increases), the advantages of doing an interim analysis become less pronounced. Interim analyses done after the accrual period would not decrease the EDA but could decrease the ETSL. If all patients are accrued prior to x^* , there can be no reduction in the expected sample size by having

an interim analysis, and the reductions in expected total study length become smaller. Clearly, if all patients are accrued at once, there can be no interim analysis and the optimal design is the fixed sample design.

As the duration of accrual increases relative to x^* , the interim analysis is done earlier relative to the fixed sample duration of accrual. As the duration of accrual approaches infinity, the optimal design becomes the Simon design (or, in our case, the normal approximation to the Simon design) since it becomes increasingly likely that each patient reaches his or her end point before the next patient is accrued. Designs optimized to minimize the EDA have their interim analyses earlier than the respective designs optimized to minimize the ETSL. This difference becomes

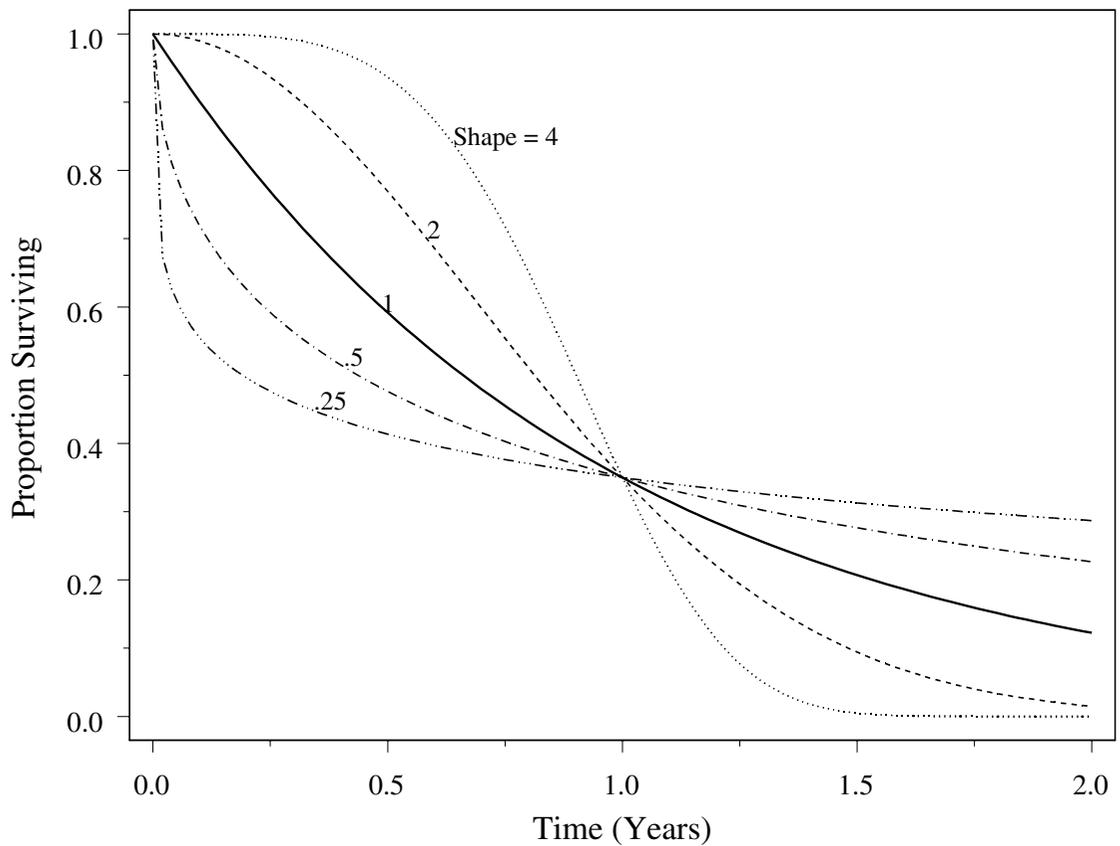


Figure 2
Weibull survival curves under the null hypothesis

Table 3: Characteristics of Optimal (for ETSL) Design Assuming Misspecification of the Survival Distributions

Weibull Shape Parameter	α	1- β
0.25	.106	.918
0.50	.103	.910
0.75	.101	.904
1.00	.100	.900
2.00	.097	.889
3.00	.096	.884
4.00	.095	.880

Table 4: Characteristics of Optimal (for ETSL) Design Assuming Misspecification of Accrual

Actual/Anticipated Accrual	Scenario 1*		Scenario 2*		Scenario 3*	
	α	1- β	α	1 - β	α	1 - β
0.25	.1	.512	.070	.687	.109	.924
0.50	.1	.711	.082	.800	.106	.918
1.00	.1	.900	.100	.900	.100	.900
1.50	.1	.963	.114	.934	.092	.863
2.00	.1	.986	.115	.970	.079	.779

* Scenario 1 – Analysis at t_1 and *MTSL* ; Scenario 2 – Analysis at t_1 and one year after t_1 or one year after accruing n patients, whichever is later; Scenario 3 – Analysis after accrual of n_1 patients and one year after accruing n patients.

Table 5: Optimal two-stage parameters for designs that minimize either the ETSL (top line) or the EDA (bottom line) for testing $H_0: S(1) \leq .35$. Power is specified for $S(1) = .5$.#

α	1 - β	DA/x^*	t_1	C_1	C_2	I_1/I_{max}	EDA	MDA	ETSL	MTSL		
.1	.9	1.50	1.01	.488	1.154	.49	1.05	1.15	.76	1.09		
			.81	-.489	1.257	.32	.96	1.03	.85	1.02		
		1.75	.93	.457	1.159	.48	1.00	1.15	.76	1.10		
			.76	-.291	1.245	.34	.94	1.05	.82	1.03		
		2	.87	.434	1.163	.47	.96	1.15	.75	1.10		
			.72	-.179	1.237	.36	.92	1.06	.80	1.04		
		3	.74	.375	1.172	.46	.88	1.15	.75	1.11		
			.63	.004	1.220	.38	.86	1.09	.77	1.07		
		4	.67	.346	1.176	.45	.84	1.15	.75	1.12		
			.59	.073	1.212	.39	.83	1.10	.76	1.08		
		.05	.95	1.50	.99	.680	1.536	.46	1.03	1.17	.72	1.10
					.81	-.274	1.622	.32	.95	1.04	.81	1.02
1.75	.91			.648	1.541	.46	.98	1.16	.72	1.10		
	.75			-.084	1.612	.34	.92	1.06	.78	1.04		
2	.85			.624	1.545	.45	.94	1.16	.71	1.11		
	.72			.024	1.605	.35	.89	1.07	.76	1.05		
3	.72			.563	1.553	.44	.85	1.16	.71	1.12		
	.63			.202	1.591	.38	.83	1.10	.72	1.08		
4	.66			.531	1.557	.44	.80	1.15	.70	1.12		
	.59			.266	1.586	.39	.79	1.11	.71	1.09		

DA is the fixed sample duration of accrual; x^* is the survival time of interest; t_1 is the time of the first interim review; C_1 and C_2 are the cutpoints for acceptance at the first stage and rejection at the second stage; I_1 and I_{max} are the information available at the interim and final reviews; *EDA* and *MDA* are the expected and maximum duration of accrual relative to the single stage values; and *ETSL* and *MTSL* are the expected and maximum total study length relative to the single stage values. Expected values are calculated under the null hypothesis.

smaller as the duration of accrual increases. As DA/x^* increases, ρ decreases slightly for designs that minimize the *ETSL* but increases for designs that minimize the *EDA*. One observes that the maximum sample size is typically 15–16% greater than the fixed sample size, regardless of the design parameters for designs that minimize the *ETSL*. The maximum sample size is smaller and increases as DA/x^* increases for designs that minimize the *EDA*. We derived results similar to those shown in Table 5 for other choices of survival under the null and alternative hypotheses (results not shown). One notes that neither the de-

signs which minimize the *ETSL* nor the ones which minimize the *EDA* are substantially affected by the choice of survival probabilities under the null and alternative hypotheses, for constant values of DA/x^* .

Simulations

The designs given in the Table 5 were derived assuming the test statistics are asymptotically normally distributed. However, the sample sizes in most phase II trials are relatively small, especially during the interim analysis, making normal approximations suspect. We performed

Table 6: Realized α and $1 - \beta$ using simulations of the optimal designs for testing $H_0: S(1) \leq .35$. Power is specified for $S(1) = .5$.

DA/x*	Desired $\alpha = .1, 1 - \beta = .9$				Desired $\alpha = .05, 1 - \beta = .95$			
	ETSL		EDA		ETSL		EDA	
	α	$1 - \beta$	α	$1 - \beta$	α	$1 - \beta$	α	$1 - \beta$
1.50	.103	.909	.100	.907	.049	.953	.043	.948
1.75	.104	.909	.114	.918	.049	.954	.045	.951
2	.126	.919	.089	.898	.050	.954	.051	.955
3	.092	.904	.116	.918	.045	.952	.059	.961
4	.092	.903	.091	.903	.045	.954	.048	.955

simulations, using one million replications for each of the designs shown in Table 5, to see how close these designs came to desired type I and II errors. Fixed sample sizes were calculated based on the exact binomial distribution, and the sample sizes at each stage of the two-stage designs were determined by multiplying the fixed sample size by t_1 and MDA from Table 5 (which are presented there relative to the fixed sample duration of accrual). These sample sizes were increased to the nearest integers. For the example discussed above, 72 patients is the fixed sample size, so for a 3 year duration of accrual (i.e., accrual rate of 24 per year), the sample size needed at the first stage is $\text{ceil}(72 * .736) = 53$ and the maximum sample size is $\text{ceil}(72 * 1.148) = 83$. The simulations were done fixing interim and final sample sizes to these values. Random entry and survival times were generated based on the survival proportions under the null and alternative hypotheses for patients during the first and second stage (if needed), and test statistics were calculated as described earlier and compared to the critical values given in the table. Results of the simulations are shown in Table 6. The realized type I and II errors are reasonably close to the desired values, and, as expected, are closer for the designs with smaller α and β (larger sample size). It is likely that designs developed for assessing larger differences between the null and alternative hypotheses (thus requiring a smaller sample size) would not be as close. Due to the discreteness of the test statistic at the final stage, small changes in the sample size can markedly affect the actual type I and II errors.

Discussion

In many phase II cancer trials, activity might be quantified by a time to event variable such as remission-free or overall survival, and the outcome might be a survival probability such as the proportion of patients remission-free or alive at one or two years. It could be that this is the primary outcome of interest due to clinical relevance or it might be that tumor response, the typical outcome in

phase II cancer trials, is not an option as all disease may have been irradiated or removed during surgery or the trial may be done only in patients who have experienced a complete response. Although designs developed for monitoring binomial proportions or modifications of these as described by Herndon [9] could also be used for these outcomes, they are not optimal. We have shown how results described by Lin et al [21] could be used to design efficient phase II trials for monitoring survival probabilities. We presented designs that minimized either the expected duration of accrual or the expected total study length under the null hypothesis. The costs of these designs are maximum sample sizes and total study lengths that are greater than the fixed sample values. However, the maximum duration of accrual and total study length are reached fairly quickly and then decrease with increasing time to the interim analysis. By considering all possible designs, one can choose a design that is almost optimal but which has a smaller sample size and maximum total study length than that of the fully optimal design.

The frequentist approach presented here represents one possible strategy for monitoring survival probabilities in a phase II trial. Other researchers have proposed rather elegant Bayesian approaches to this problem. Follman and Albert [24] use a Dirichlet prior distribution for describing the probabilities of failure at discrete times. They show that the posterior distribution incorporating censored data is a mixture of Dirichlets, and they use simulations to estimate the posterior probability that an event rate exceeds some threshold. Cheung and Thall [25] present a method for monitoring survival probabilities based on an approximate posterior, and they apply stopping rules described in Thall and Simon [6] based on the approximate probabilities. In addition, they extend their methods to the more complicated case of composite events. Both of these methods have the advantage that they can be applied continuously.

Another approach to the design of phase II trials for monitoring survival probabilities would be to use parametric methods. The survival probability at a given time would be a function of the parameters of the parametric model. For example, assuming an exponential model, monitoring a survival probability at a given time is equivalent to monitoring the hazard parameter. Methods proposed by Case and Morgan [26] could then be used to obtain optimal durations of accrual and follow-up. This approach has the advantage of using events that occur before or after the time of interest, but the disadvantage of relying on a specific model.

In the designs discussed above, we assumed a parametric distribution to obtain estimates of the information over time. Were we to have access to data or simply the Kaplan-Meier estimate of the survival distribution from a previous trial, we could use those data to obtain a nonparametric estimate of the information using equation (3). In that equation, the denominator would need to be multiplied by $\min[(t - X_i)/MDA, 1]$. This fully nonparametric approach has the conceptual advantage that when we design a study to compare nonparametric estimates of survival distributions, we frequently do not want to make assumptions about the form of $S(\cdot)$. Although one would like to know $S(\cdot)$ with certainty in planning a study, often we are faced with an estimate of a summary statistic such as a percentile, median, or mean and maybe an idea of the shape from a Kaplan-Meier plot without access to the data. Sometimes we have the data from a previous study but do not have confidence that the population is relevant to the planned study. Often the choice of developing a design using the parametric or nonparametric approach will depend on the existence of previous raw data and, if such data exist, our relative confidence in the parametric assumptions versus the applicability of the population in the previous study.

Although the approach described above is applicable when the duration of accrual is long relative to the survival time of interest, it is not practical when the duration of accrual is short (say less than 1.5 times the survival time of interest). This is because there is little information available for estimating the survival probability before most of the patients are accrued to the study. If the accrual period is less than the time of interest, there is no design that leads to an expected duration of accrual that is less than the fixed sample value. In our experience of doing phase II trials in a single institution, it has been rare that fast accrual would have limited the applicability of these designs. However, it could be more of a problem with multicenter or cooperative group trials. A possible solution includes monitoring a different (earlier) time during the interim analysis as Lin et al [21] did in their example.

Unfortunately, the connection with the primary end point of interest is not always clear.

Misspecification of the accrual rate can have a major effect on the operating characteristics of these designs. The misspecification of survival during the design phase has much less impact on the type I and II errors. However, it is clear that one needs to monitor the ongoing accrual (as is typically done in clinical trials) and make modifications to the design once the accrual rate is ascertained. The ultimate critical values can be modified once the actual estimate of the information is obtained during the interim analysis, but this strategy should result in fairly efficient designs.

Competing interests

None declared.

Authors' contributions

Both authors contributed to all sections of this work; both read and approved the final manuscript.

Acknowledgements

Supported in part by grants P30-CA-12127 and U10-CA-81851 from the Public Health Service, National Institutes of Health.

References

- Gehan EA **The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent** *Journal of Chronic Diseases* 1961, **13**:346-353
- Fleming TR **One sample multiple testing procedure for phase II clinical trials** *Biometrics* 1982, **38**:143-151
- Chang MN, Therneau TM, Wieand HS and Cha SS **Designs for group sequential phase II clinical trials** *Biometrics* 1987, **43**:865-874
- Simon R **Optimal two-stage designs for phase II clinical trials** *Controlled Clinical Trials* 1989, **10**:1-10
- Green SJ and Dahlberg S **Planned versus attained design in phase II clinical trials** *Statistics in Medicine* 1992, **11**:853-862
- Thall PF and Richard Simon **Practical Bayesian guidelines for phase IIB clinical trials** *Biometrics* 1994, **50**:337-349
- Chen TT **Optimal three-stage designs for phase II cancer clinical trials** *Statistics in Medicine* 1997, **16**:2701-2711
- Heitjan DF **Bayesian interim analysis of phase II cancer clinical trials** *Statistics in Medicine* 1997, **16**:1791-1802
- Herndon JE II **A design alternative for two-stage, phase II, multicenter cancer clinical trials** *Controlled Clinical Trials* 1998, **19**:440-450
- Chen TT and Ng T-H **Optimal flexible designs in phase II clinical trials** *Statistics in Medicine* 1998, **17**:2301-2312
- Colton T and McPherson K **Two-stage plans compared with fixed-sample-size and Wald SPRT plans** *J Am Stat Assoc* 1976, **71**:80-86
- Landis SH, Murray T, Bolden S and Wingo PA **Cancer Statistics, 1998** *CA: A Cancer Journal for Clinicians* 1998, **48**:6-29
- Lawrence T **Gemcitabine as a radiation sensitizer** *Sem Oncol* 1995, **22**:68-71
- Casper ES, Green MR, Kelsen DP, Heelan RT, Brown TD, Flombaum CD, Trochanowski B and Tarassoff PG **Phase II trial of gemcitabine (2',2'-difluoro-2'-deoxycytidine) in patients with adenocarcinoma of the pancreas** *Investigational New Drugs* 1994, **12**:29-34
- Burriss HA 3rd, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, Cripps MC, Portenoy RK, Storniolo AM, Tarassoff R, Nelson R, Dorr FA, Stephens CD and Von Hoff DD **Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial** *J Clin Oncol* 1997, **15**:2403-2413

16. Blackstock AW and Bernard SA **Twice weekly gemcitabine and concurrent radiation: laboratory studies supporting phase I clinical trials in pancreatic cancer** *Can Conf* 1999, **3**:2-6
17. Kaplan EL and Meier P **Nonparametric estimation from incomplete observations** *J Am Stat Assoc* 1958, **53**:457-481
18. Nelson W **Hazard plotting for incomplete failure data** *J Quality Technology* 1969, **1**:27-52
19. Jennison C and Turnbull BW **Group Sequential Designs with Applications to Clinical Trials** Boca Raton, Chapman & Hall/CRC 2000.
20. Lan KKG and DeMets DL **Discrete sequential boundaries for clinical trials** *Biometrika* 1983, **70**:659-663
21. Lin DY, Shen L, Ying Z and Breslow NE **Group sequential designs for monitoring survival probabilities** *Biometrics* 1996, **52**:1033-1042
22. Donnelly TG **Algorithm 462 – Bivariate normal distribution [S15]** *Comm ACM* 1973, **16**:638
23. Brent R **Algorithms for Minimization without Derivatives** Englewood Cliffs, Prentice Hall 1973,
24. Follmann DA and Albert PS **Bayesian monitoring of event rates with censored data** *Biometrics* 1999, **55**:603-607
25. Cheung YK and Thall PF **Monitoring the rates of composite events with censored data in phase II clinical trials** *Biometrics* 2002, **58**:89-97
26. Case LD and Morgan TM **Duration of accrual and follow-up for two-stage clinical trials** *Lifetime Data Analysis* 2001, **7**:21-37

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/3/6/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

